Tilted Empirical Risk Minimization  
Tian Li* (CMU), Ahmad Beirami* (Facebook AI), Maziar Sanjabi (Facebook AI), Virginia Smith (CMU)

**Tilted ERM (TERM) Objective**

Address the deficiencies of ERM $\min R(\theta) := \frac{1}{n} \sum_{i=1}^{n} f(x_i, \theta)$ in a unified framework:

$\text{Tilted ERM: } \min R_t(\theta) := \frac{1}{n} \sum_{i=1}^{n} e^{\theta^T f(x_i, \theta)}$  

Use the tilt parameter $t$ to tune the impact of individual losses.

**Properties**

- Re-weighting samples to magnify/suppress outliers
- $R_t(\theta) := \frac{1}{n} \log \left( \frac{1}{n} \sum_{i=1}^{n} e^{\theta^T f(x_i, \theta)} \right)$  

$\text{grads: } \nabla_{\theta} R_t = \sum_{i=1}^{n} e^{\theta^T f(x_i, \theta)} f_i(x_i, \theta)$  

Trade-off between average loss and max/min-loss

- As $t$ moves from 0 to $\infty$, the average loss will increase, and the max-loss will decrease.
- As $t$ moves from 0 to $-\infty$, the average loss will increase, and the min-loss will decrease.

**Approximation of quantile losses**

$\tilde{q}_t \approx \min_{\theta} \sum_{i=1}^{n} \mathbb{1}(f(x_i, \theta) \geq t)$

TERM objectives for a squared loss problem with $N=3$. Tilted losses recover min-loss, avg-loss, and max-loss. TERM is smooth for all finite $t$ and convex for positive $t$. TERM is competitive to state-of-the-art methods for classification with imbalanced classes.

**Applications**

- On real-world ML applications, TERM is superior to (or competitive with) existing, problem-specific state-of-the-art solutions.
- Robust regression
- Robust classification
- Other applications of the TERM framework (e.g., meta-learning, GAN training)
- Other properties of TERM (e.g., adversarial robustness)
- Generalization of the TERM objective with respect to $t$
- Further connections with other risks (DRO, Conditional Value-at-Risk, Invariant Risk Minimization, etc.)

**Future Work**

- TERM is competitive with robust regression baselines, particularly in high noise regimes.
- TERM completely removes the impact of noisy annotations.
- TERM is able to handle compound issues, e.g., the existence of noisy samples and imbalanced classes.

**Code:** https://github.com/litian96/TERM